THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 8, October 1977
Printed in U.S.A.

An Evaluation of Two Simple Methods for Detecting Tones over Telephone Lines

By D. K. CHRISTOPHER, L. R. RABINER, P. SCHWEITZER, and D. E. BOCK

(Manuscript received March 24, 1977)

An important practical application of signal processing theory is the problem of complex tone detection. Within the telephone plant there often arises a need for a simple, yet efficient, method for detecting the various tones which are used in telephone communication. Two such methods are discussed in this paper. One method uses measurements of the short-time signal energy and makes the decision as to whether or not there is a particular tone present on the line based on the periodicity of the envelope of the signal. This method has application in determining if the energy on the line is periodic or aperiodic where sample examination time is not limited. The second method uses measurements of the short-time zero crossings of the signal. A parallel processing scheme is used to determine if a particular tone is present based on the detailed statistical properties of each of the tones. This method has application in determining if specific frequencies are present, especially when the examination time of the sample is limited. Using a large number of dialed-up connections, both systems were evaluated as to accuracy and speed. Results are presented which show the properties of the two tone detection methods.

I. INTRODUCTION

The need to reliably detect tones arises in a number of systems which are in use within the telephone plant. A wide variety of methods have been proposed for solving this problem including digital filtering, spectral analysis etc.[1–4] In this paper we discuss a particular problem in tone detection and show some characteristics of two simple systems designed to solve this problem.

Figure 1 shows a pictorial description of a simple telephone call in which the calling party initiates a call and the call is switched through a central office. When the called party picks up the telephone a dc path

Fig. 1—Pictorial description of a simple telephone call.

is completed to the central office to indicate that the connection has been made, and that billing of the calling party can begin. The indication is formally called answer supervision. On a certain percentage of calls, anomalies occur such that calls are not completed in the normal manner. A recent survey of a total of 3 million unanswered calls indicated that 90 seconds after the calling party initiated the call, no answer supervision had been received on 61,000 (approximately 2 percent) of the calls. There are several possibilities which account for these seemingly long unanswered calls. These include:

(*i*) Persistent callers—i.e., the calling party is waiting 90 seconds (15 ring cycles) for the called party to answer the telephone. Another possibility is that the line is busy, and the calling party remains on the line in spite of hearing 90 seconds of busy tone. Yet another possibility is that the call was improperly routed and that the calling party is listening to a fast busy or reorder signal.

(*ii*) Announcement service—e.g., the calling party dialed an inoperative number and is listening to an announcement concerning the called telephone. Generally such announcements are short and will not last 90 seconds, but it is not impossible for this to occur.

(*iii*) Defective telephone—i.e., the telephone of the called party is defective or the circuitry which generates the answer supervision signal is not working.

(*iv*) Network irregularities—i.e., the equipment making up the telephone network interconnecting telephone offices may be defective or inoperative and fail to relay the proper answer supervisory signals.

It is important to be able to isolate cases (*iii*) and (*iv*) in order to take appropriate action. Since direct methods for determining whether a telephone is defective or whether a network irregularity exists are not easily implemented, an indirect approach is indicated. The approach studied here was one in which the signal on the line was processed to determine whether or not tones were on the line [case (*i*)], and if not, whether speech could be detected [cases (*ii*)–(*iv*)].

Although sophisticated approaches to tone detection are applicable, such methods are unnecessarily complex (and expensive) for the problem of detecting the three tones described above. Thus two relatively simple, and yet very different, methods for detecting tones were studied. The first method was developed on the assumption tbat the frequencies of

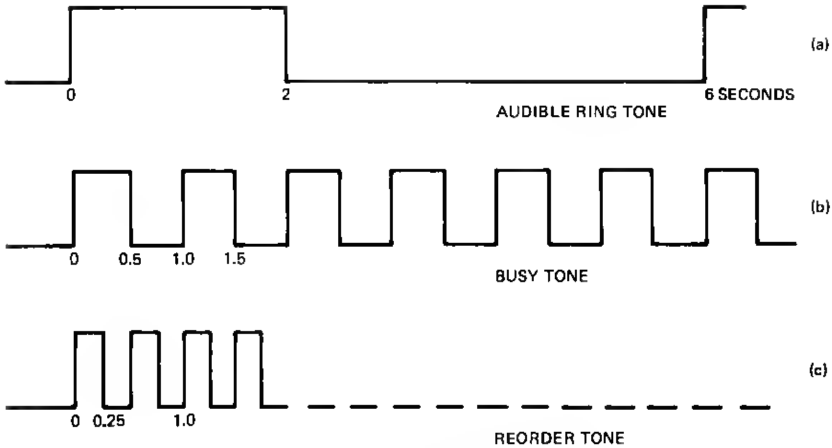

Fig. 2—Temporal characteristics of the tones.

the individual tones would vary greatly from one central office to another and that gathering adequate statistics on these frequencies was impractical. Therefore long-term properties of the tones were used in the detection algorithm. This first approach, known herein as the energy system, was developed by F. T. Boesch and R. E. Thomas[4] to comply with the above constraint. The second method uses measurements of the zero crossing rate of the signal, and used a parallel processing scheme to determine if a particular tone is present based on the assumption that detailed statistical properties were available for each of the tones. This method was developed in the Acoustics Research Department. Descriptions of each of these two methods are given in Sections III and IV. In the next section we discuss the properties of the individual tones for which the two systems were designed. Finally, in Section V we give the results of simulation experiments with both tone-detection methods.

II. PROPERTIES OF THE TONES

The three tones which had to be detected were: (*i*) audible ring tone, (*ii*) line busy tone, (*iii*) reorder tone. Figure 2 shows a sketch of the nominal temporal pattern of these three tones. The audible ring tone is on for 2 seconds and off for 4 seconds. The line busy tone is on for 0.5 second, and off for 0.5 second. The reorder (fast busy) tone is somewhat variable in its temporal pattern. Generally its overall period is 0.5 second, with an on period of from 0.2 to 0.3 second, and an off period from 0.3 to 0.2 second. Nominally the on and off periods are 0.25 second.

Within one period the spectral properties of these tones can vary a fair amount. This is due both to the variability in the mechanical equipment which produces the tones (e.g., motor generators), and to the different

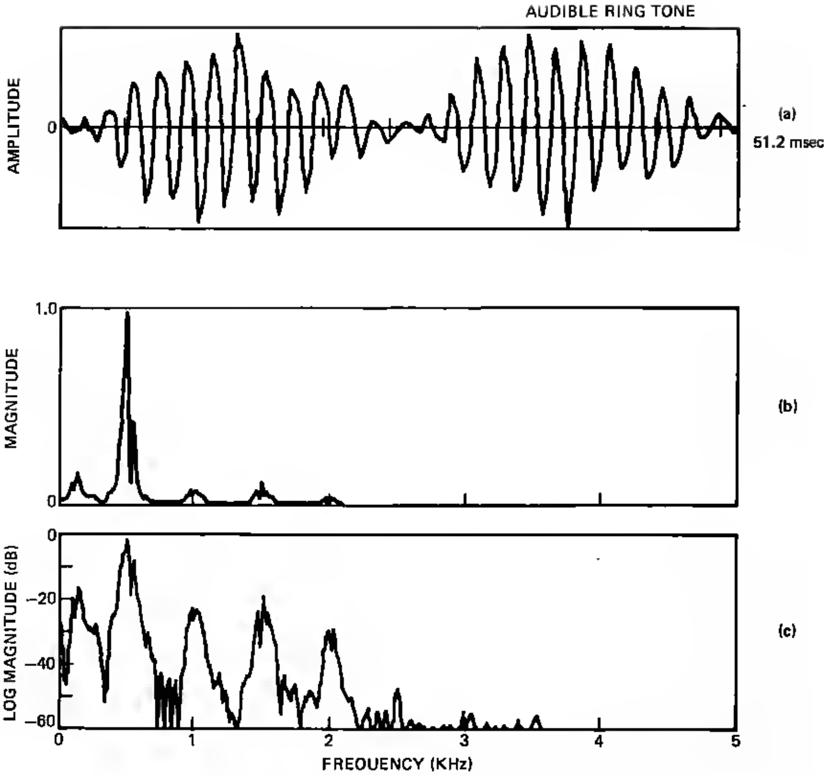

Fig. 3—Spectral characteristics of audible ring tone.

standards in use within the telephone plant. However, the dominant amount of spectral energy in these tones is concentrated in a region around 500 Hz. By way of example, Figs. 3 to 5 show plots of typical 51.2 msec sections of audible ring (Fig. 3), busy tone (Fig. 4), and reorder tone (Fig. 5), along with linear and log magnitude spectra for these tones. A Hamming window was used on the data to minimize the effects of the endpoints of the signal on the resulting spectrum. It can readily be seen from these figures that these tones have a reasonably complex structure both in time and in frequency.

Preliminary analysis also uncovered some prominent temporal properties of these tones with which the tone-detection systems would have to deal. For the audible ring tone a substantial transient generally was found at the beginning and end of each on cycle. Figure 6 shows an example of such a transient occurring at the beginning of a cycle. Such transients are distinctly audible as clicks. Also for each of the tones a substantial amount of variability in the duration of the on-off cycles was also observed—even within consecutive cycles of the same tone. Figure

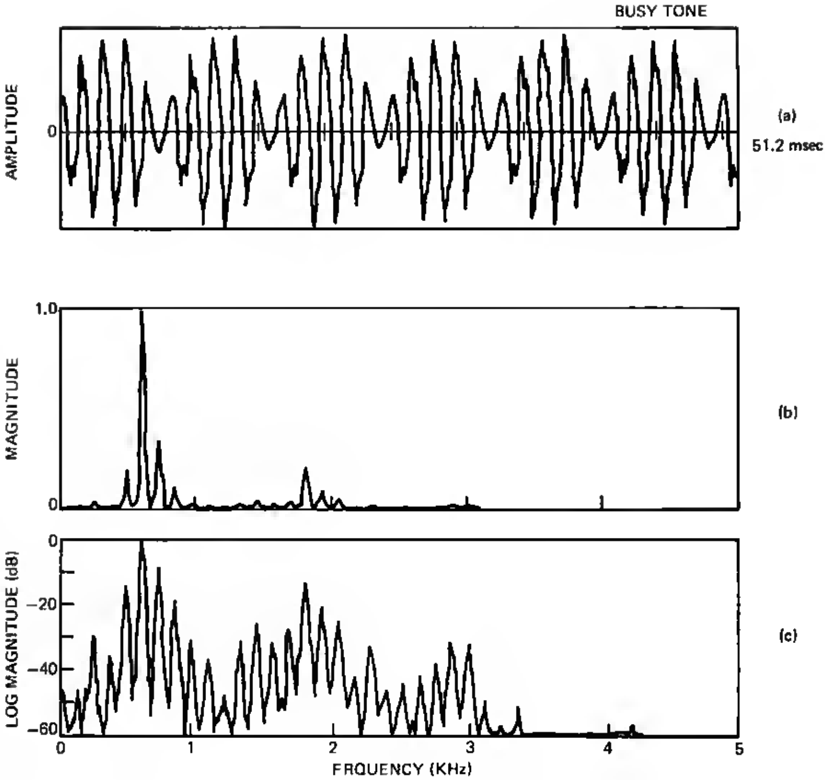

Fig. 4—Spectral characteristics of busy tone.

7 illustrates this effect during 2 cycles of a busy tone. The first on cycle lasts about 0.44 second, whereas the second on cycle lasts about 0.47 second. Such variability was not uncommon for the tones which were studied.

III. ENERGY-BASED SYSTEM FOR TONE DETECTION

Figure 8 shows a block diagram of the simulation of the energy-based system for tone detection. The input signal $x(t)$ is first band-pass filtered to the range 300–3200 Hz, and then sampled at a 10-kHz rate. An energy contour of $x(n)$ is computed using a 501-point (50-msec) Hamming window to give

$$E(n) = \sum_{m=0}^{N-1} [x(n-m)w(m)]^2 \tag{1}$$

where

$$N = 501$$

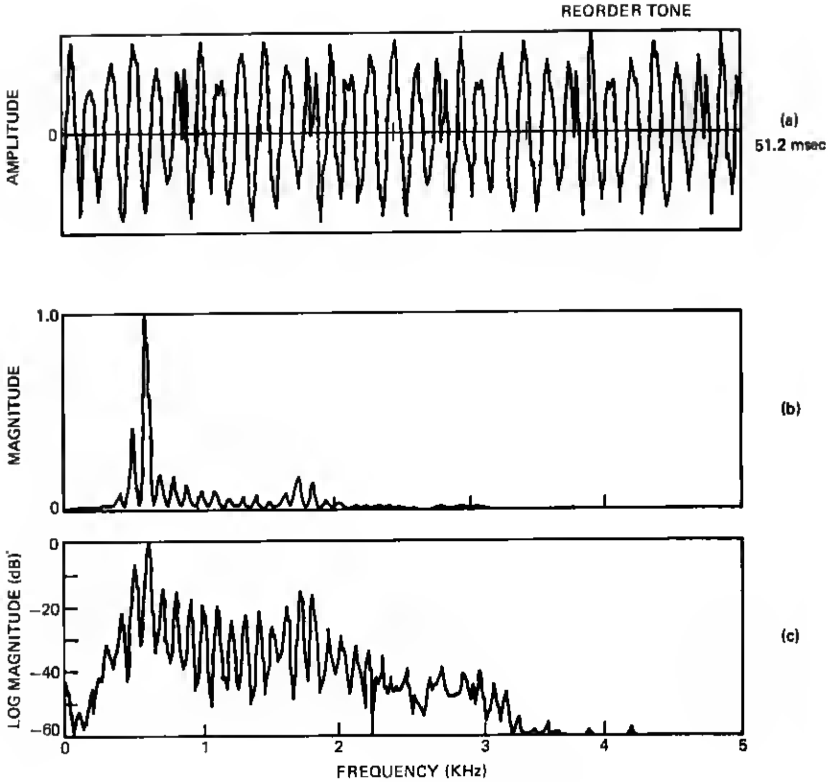

Fig. 5—Spectral characteristics of reorder tone.

and

$$w(n) = 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1}\right) \tag{2}$$

The energy contour is resampled at a rate of 20 times/sec.* A noise threshold is computed by finding the minimum energy of the signal, and setting the threshold 12 dB above this level. Based on the energy threshold, the signal $E(n)$ (at the 20-Hz rate) is infinite-peak-clipped to give a binary signal $b(n)$, which is of the form

$$\begin{aligned} b(n) &= 1 \quad \text{if } E(n) \geqslant T \\ &= 0 \quad \text{if } E(n) < T \end{aligned} \tag{3}$$

(To eliminate the effects of spurious transient on the line, a delay of 150 msec, i.e., 3 samples, is built into the infinite peak clipper for off-on

* In the implementation $E(n)$ is computed only 20 times/second.

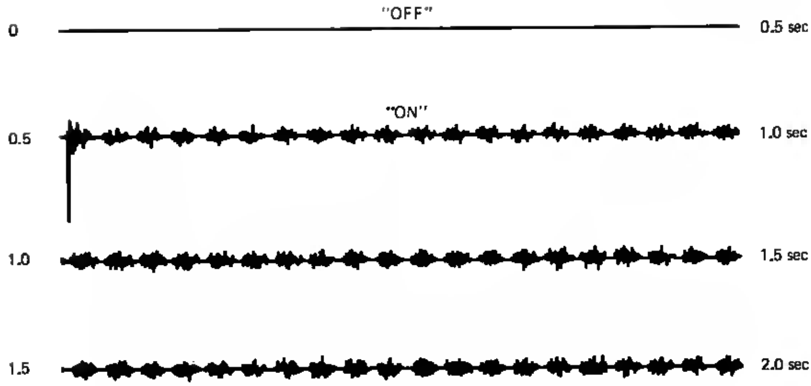

Fig. 6—Example showing transient at beginning of audible ring tone.

transitions to guarantee that $E(n)$ stays above T for this period. If $E(n)$ falls below T during this period, then $b(n)$ stays at 0. Similarly a delay of 50 msec, i.e., 1 sample, is built into on-off transitions. Thus, $E(n)$ must fall below T for 2 consecutive samples for $b(n)$ to be set to 0).

Following clipping, the signal $b(n)$ is blocked into runs. A run is defined as a sequence of 1's followed by a sequence of 0's. The detection system processes $x(n)$ until a total of 5 runs is obtained, or until 40 seconds of data are processed, whichever occurs first. The duration of each run, $r(j)$, $j = 1,2,\ldots 5$, is measured, and the average run length, RL, is determined as

$$\text{RL} = \frac{1}{5}\sum_{j=1}^{5} r(j) \tag{4}$$

The signal, $b(n)$, is then comb-filtered using a fixed comb of delay RL samples, giving

$$c(n) = b(n) - b(n - RL) \tag{5}$$

The signal $c(n)$ is of the form

$$\begin{aligned} c(n) &= +1 \quad \text{if } b(n) = 1,\ b(n - RL) = 0 \\ &= 0 \qquad \text{if } b(n) = 0,\ b(n - RL) = 0 \\ &\qquad\quad\ \ \text{if } b(n) = 1,\ b(n - RL) = 1 \\ &= -1 \quad \text{if } b(n) = 0,\ b(n - RL) = 1 \end{aligned} \tag{6}$$

The absolute value of $c(n)$ is then accumulated, and the result is normalized by dividing by the number of samples of $b(n)$ which went into the computation, giving

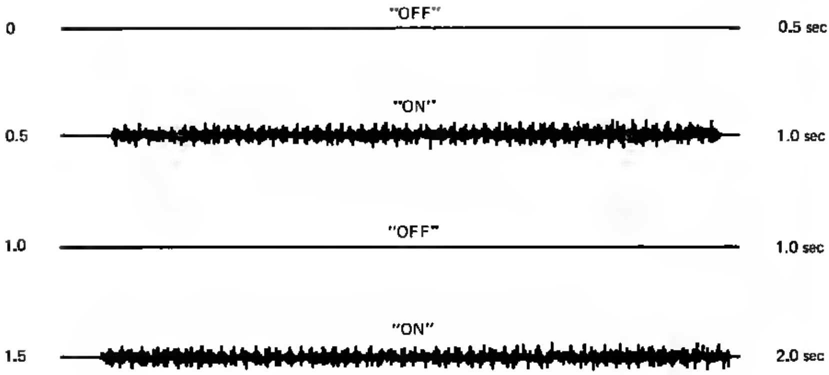

Fig. 7—Example showing variation in on-off cycle of busy tone.

$$D_N = \frac{\sum_{n=0}^{N-1} |c(n)|}{N} \tag{7}$$

Clearly D_N is a normalized measure of the lack of periodicity of the signal, since

$$0 \leqslant D_N \leqslant 1 \tag{8}$$

and $D_N \rightarrow 0$ if the signal is perfectly periodic and of period RL.

The final tone detection is based on the values of RL and D_N. If D_N is sufficiently small (indicating the signal is periodic) then the signal is classified as a tone of period RL samples. The tone whose period is closest to RL samples is chosen as the correct tone. If D_N is sufficiently large (indicating a lack of signal periodicity) then the signal is classified as speech (or silence). Based on experimentation with the system the thresholds chosen for D_N and the corresponding decision rules are

if $D_N < 0.15$	signal is periodic (tone) of period R_L samples
$D_N > 0.30$	signal is aperiodic (speech or silence)
$0.30 > D_N > 0.15$	signal is undefined

The undefined region accounts for tones on a very noisy-line, or speech conversations with a fairly periodic rhythm of talking.

Figure 9 shows some typical energy contours $[E(n)]$ for 15-second recordings of an audible ring tone (Fig. 9a), a reorder tone (Fig. 9b), a busy tone (Fig. 9c), and a weather announcement (Fig. 9d). It can be seen that by appropriate placement of the silence energy threshold, the resulting binary signal $b(n)$ will be periodic. However, a variation of over

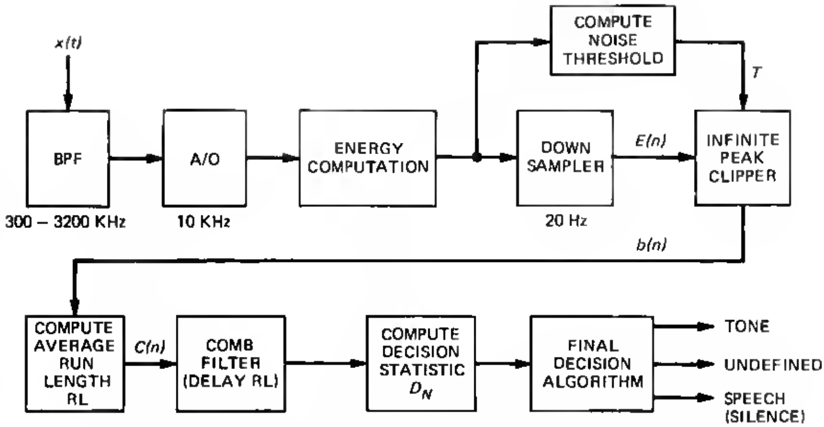

Fig. 8—Block diagram of energy system for tone detection.

20 dB in the level of $E(n)$ is obtained between the audible ring and the reorder signal. Thus careful choice of the silence threshold is extremely important to the proper functioning of this system. The transients present in the audible ring signal are also clearly seen in Fig. 9a, as noted previously.

IV. ZERO-CROSSING-BASED SYSTEM FOR TONE DETECTION

Figure 10 shows a block diagram of the zero crossing system which was used for tone detection. This system is organized as a parallel processing system with an individual detector for each tone. Speech (or silence) is indicated by the absence of a detected tone for an 8-second interval. The operation of this system is as follows. The input signal, $x(n)$, is sampled at a 10-kHz rate and then fed into three parallel tone detectors. The output of the tone detector is zero until a tone is detected at which point the output becomes one until the detector is reset by the decision logic. The decision logic to choose the tone is very simple. The output of each tone detector is monitored at a fixed rate until either one of the lines indicates a tone, or until 8 seconds have passed.* If one of the tone lines indicates a tone the logic decides which tone was detected and resets the tone detectors. After 8 seconds without a tone indication, the logic classifies the signal as speech (or silence).

Figure 11 shows a block diagram of the individual tone detectors. For each tone detector there are two parameters which define the range of the level crossing parameter for a particular tone. A level crossing for the signal $x(n)$ occurs at $n = n_0$ when

* The choice of a maximum interval of 8 seconds is explained later.

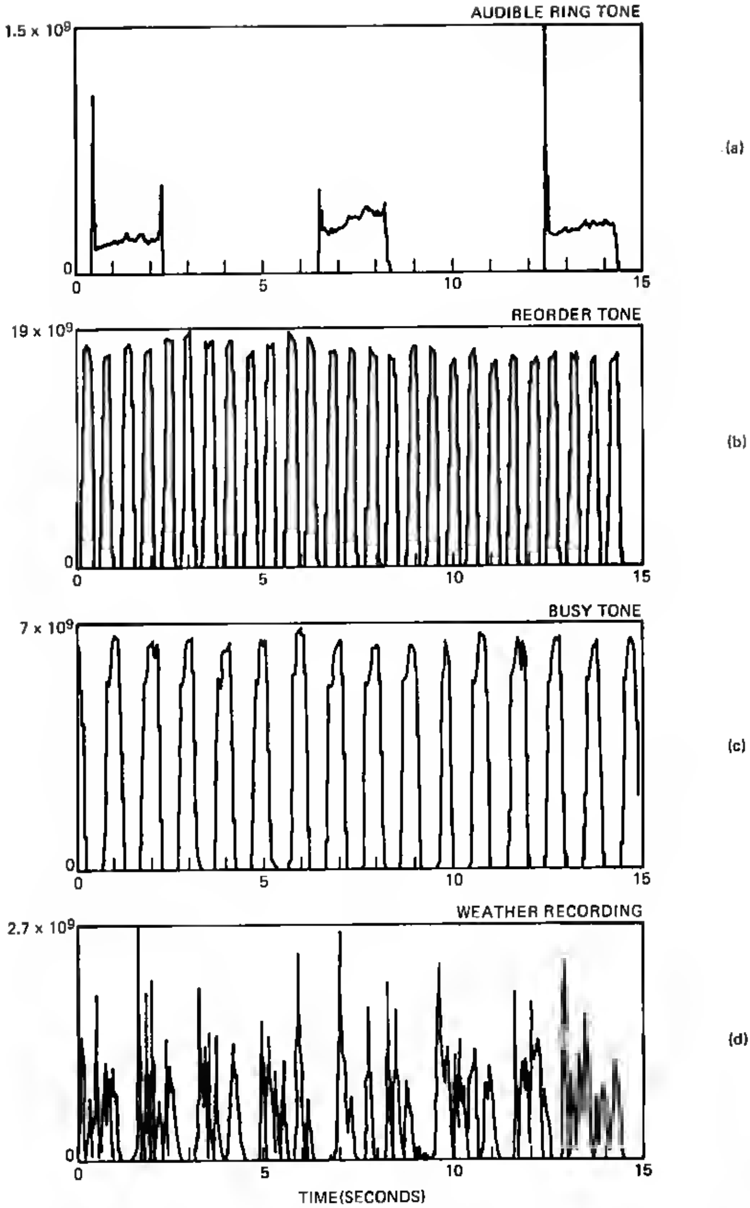

Fig. 9—Typical energy contours of three tones and a weather announcement.

$$x(n_0) \geqslant T \tag{9}$$

and

$$x(n) < -T \tag{10}$$

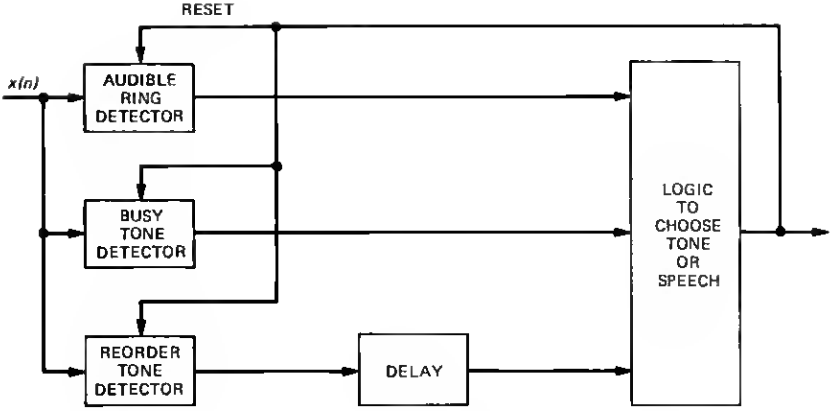

Fig. 10—Block diagram of the zero crossing system for tone detection.

for some value of $n = n_1 < n_0$ where n_1 is greater than the value at which the previous level crossing occurred—i.e., the signal must have been below the level $(-T)$ and risen above the level T for a level crossing to have occurred. The parameter used for tone detection was the number of level crossings of the signal within a specified duration D. Call this parameter $L_x(T,D)$. For each D second duration the quantity $L_x(T,D)$ is computed and compared to the range parameters R_L and R_H which define an expected range for $L_x(T,D)$ for the jth tone. For the jth tone to be detected the parameter $L_x(T,D)$ must satisfy the relation

$$R_L \leqslant L_x(T,D) \leqslant R_H \tag{11}$$

for three consecutive D second intervals. This sequence is called a triple

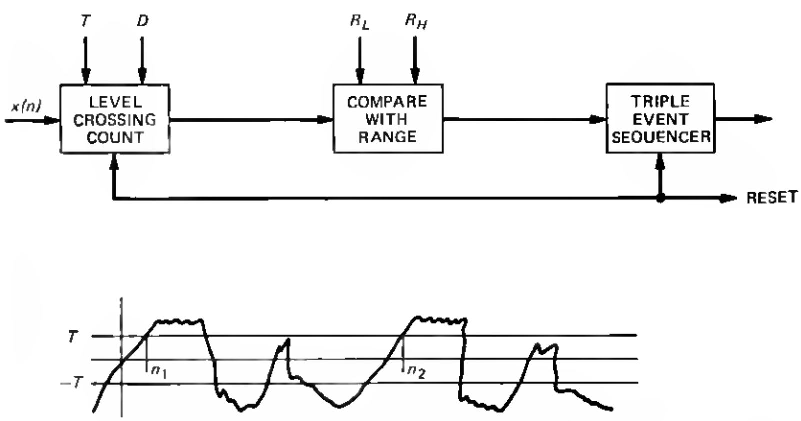

Fig. 11—Block diagram of the individual tone detectors.

Table I — Values for three tones

Signal	D	T	R_L	R_H
Reorder tone	60 msec	200	34	37
Busy tone	120 msec	200	67	73
Audible ring tone 1	480 msec	100	190	245
Audible ring tone 2	480 msec	100	220	270

event—an event being whenever eq. (11) is satisfied. The reason a triple event sequencer is used is because the major concentration of energy for the tones being detected is in the range 200–1000 Hz, and this in tbe region in which the major concentration of speech energy also occurs. Thus it is quite possible for eq. (3) to be satisfied by a speech signal, as well as by the correct tone. To minimize the possibility of a speech signal being detected by the tone logic, the triple event sequencer was used.

Table I gives the values of T,D, R_L and R_H for the three tones which were used in this investigation. For each tone the quantity D was chosen to be approximately ¼ of the on cycle of the tone to guarantee that at least 3 complete cycles of D seconds would occur within the on period of the tone, independent of the phase of the initial D second region. (Recall that the D second intervals are asynchronous with the tone—thus an interval may contain a transition from on to off or vice versa.) In addition, for both software and hardware convenience, the values of D are all multiples of 60 msec.

It is seen in Table 1 that two sets of parameters were used for the audible ring tones. This was due to the bimodal distribution of $L_x(T,D)$ which was measured when representative tones were recorded in different areas. In practice, an additional tone detector is used for such cases.

The reason for the delay in the reorder tone detector of Fig. 10 should now be clear. Since the frequency characteristics of the busy tone and the reorder tone were almost identical, a busy tone on the line would cause the reorder tone detector to detect the tone before the busy tone detector. Thus, a delay of 240 msec was used to allow the busy tone detector a chance to detect a busy tone prior to classifying the tone as a reorder tone.

The purpose of the reset signal following each signal classification is to clear the level crossing count, and to clear the triple event sequencer, so that the overall tone detector can be switched to a new line in order to classify a new signal.

Finally the reason for waiting 8 seconds until classifying the signal as speech (or silence) is that in the worst case the tone detector might be switched into an audible ring tone just past the beginning of an on cycle. Thus, the detector would not indicate the audible ring tone during this first on cycle, and would have to wait for the second on cycle for detecting

Table II — Results of tests on audible ring tone

(a) Accuracy of detection
Signal—Audible Ring Tone

Connection	Energy System	Zero Crossing System
1	100/100	100/100
2	100/100	100/100
3	100/100	100/100
4	100/100	100/100
5	100/100	100/100
6	100/100	100/100
Total	600/600	600/600
Percentage Accuracy	100%	100%

(b) Speed of detection

Connection	Energy System	Zero Crossing System
1	33.70 sec	4.574 sec
2	33.83	4.291
3	31.70	4.142
4	33.77	4.569
5	33.22	3.912
6	33.32	4.032
Total	189.54 sec	25.518 sec

$$\text{Relative Time} = \frac{189.54}{25.518} = 7.43$$

the tone. This would require a total of 2 seconds for the first on cycle, plus 6 seconds for the second off-on cycle, or a total of 8 seconds before audible ring can be eliminated.

V. EXPERIMENTAL EVALUATIONS

An extensive experimental evaluation of these two methods for tone detection was carried out over standard dialed-up telephone lines. For each type of signal (i.e., the three tones, and speech) a number of different connections was tested. For each connection a total of 60 seconds of the signal was recorded. A total of 100 trials were made on each signal with each trial beginning at a randomly selected point in the recording.

Results of these evaluation tests are given in Tables II to V for audible ring, busy, reorder, and speech respectively. Each table shows the accuracy for each system, as well as the average length of signal required to make the decision. Thus, for the audible ring tone (Table II), six different telephone numbers were used giving a total of 600 trials. Both systems detected audible ring 100 percent reliably. However, the energy-based system required about 7.4 times the amount of signal required by the zero crossing system—i.e., the average time to detect audible ring tone was 4.4 seconds for the zero crossing system, and 31.9 seconds for the energy system. (It should be kept in mind that the two tone-detection

Table III — Results of tests on busy tone

(a) Accuracy of detection
Signal—Busy Tone

Connection	Energy System	Zero Crossing System
1	100/100	79/100 (21 Ro)
2	100/100	77/100 (23 Ro)
3	99/99	82/100 (18 Ro)
4	100/100	91/100 (9 Ro)
5	100/100	79/100 (21 Ro)
6	100/100	84/100 (16 Ro)
7	100/100	85/100 (15 Ro)
Total	699/699	577/700 (123 Ro)
Percentage Accuracy	100%	82.4% → 100%

(b) Speed of detection

Connection	Energy System	Zero Crossing System
1	5.77 sec	0.596 sec
2	5.81	0.592
3	5.78	0.666
4	5.65	0.643
5	5.84	0.560
6	5.74	0.614
7	6.09	0.678
Total	40.48 sec	4.349 sec

Relative time $= \frac{40.48}{4.349} = 9.31$

systems were designed with different constraints, as discussed previously.)

For the busy tone (Table III) a total of seven different connections, or 700 trials, were used. The energy-based system detected 699 out of 699 correctly,* whereas the zero crossing system detected only 577 out of 700 correctly. Of the 123 errors, all were classified as reorder signal. This is in fact no real error since it occurs whenever the signal starting point is past the beginning of an on cycle, but before the middle of the on cycle—a fairly common occurrence. In such cases the busy tone is indistinguishable from a reorder tone. Thus the zero crossing system was effectively 100 percent accurate.

In terms of processing time the zero crossing system took, on average, 0.62 second to detect busy tone, whereas the energy system took about 5.8 seconds. Thus the zero crossing system was a factor of 9.3 times faster than the energy system.

Table IV shows the results for the reorder tone tests. A total of five connections were tested resulting in 500 individual trials. The energy system accurately detected 493 of 500, or 98.6 percent of the trials. The 7 errors involved classifying the signal as speech. These errors were

* On one trial the random starting point was too close to the end of the recording so no decision was made.

Table IV — Results of tests on reorder tone

(a) Accuracy of detection
Signal—Reorder Tone

Connection	Energy System	Zero Crossing System
1	100/100	100/100
2	100/100	100/100
3	100/100	100/100
4	93/100	100/100
5	100/100	87/100
Total	493/500	487/500
Percentage Accuracy	98.6%	97.4%

(b) Speed of detection

Connection	Energy System	Zero Crossing System
1	2.98 sec	0.576 sec
2	3.00	0.572
3	3.18	0.572
4	3.35	0.583
5	2.99	1.526
Total	15.40 sec	3.829 sec

$$\text{Time Ratio} = \frac{15.4}{3.829} = 4.02$$

eliminated by raising the noise level threshold to a value 18 dB above the noise level (instead of 12 dB as was normally the case). For the zero crossing system the accuracy was 487 of 500 or 97.45 percent of the trials. Of the 13 errors, 2 were classified as audible ring tone, and 11 were classified as speech. All these errors were corrected by increasing the search duration to include a second cycle of the tone.

In terms of speed the energy system required, on average, 3.08 seconds to detect the reorder tone, whereas the zero crossing system required about 0.76 second. Thus, the zero crossing system was a factor of 4 times faster than the energy system for this tone.

For testing the two systems on speech signals, two classes of signals were used. One class was of the announcement type—i.e., weather, news service, recorded phone messages, etc. The other class was a set of conversions. Table V shows the results on these two sets of signals. For recording and announcements the energy system detected 228 of 272 trials, or 83.8 percent of the trials. (In 128 cases the random starting point of the message was too close to the end of the message for a decision to have been made). All 44 errors were cases when the threshold fell in the undefined region. For conversational speech the energy system detected 398 of 398 trials, for 100 percent accuracy.

For the zero crossing system on recorded speech an accuracy of 364 of 400 trials, or 96 percent was obtained, and for conversations an accuracy of 390 of 400 trials or 97.5 percent was obtained. For recorded announcements 15 of the 16 errors were audible ring tone, and in one case

Table V — Results of tests on speech

(a) Accuracy of detection
Signal-Speech Recordings and Conversation

1. Recordings

Connection	Energy System	Zero Crossing System
1–Chicago Weather	25/43–18UD	84/100
2–NYC Weather	100/100	100/100
3–Viking News	3/29–26UD	100/100
4–Recorded Phone Trouble	100/100	100/100
Total	228/272–44UD	384/400
Percentage Accuracy	83.8%	96.0%

2. Conversation

Connection	Energy System	Zero Crossing System
1–2 Females	100/100	93/100
2–1 Female–1 Male	100/100	99/100
3–1 Female–1 Male	98/98	98/100
4–1 Female–1 Male	100/100	100/100
Total	398/398	390/400
Percentage Accuracy	100%	97.5%

(b) Speed of detection

1. Recordings

Connection	Energy System	Zero Crossing System
1	25.42 sec	7.347 sec
2	12.89	8.040
3	38.04	8.040
4	6.44	8.040
Total	82.79 sec	31.467 sec

$$\text{Time Ratio} = \frac{82.79}{31.467} = 2.63$$

2. Conversation

Connection	Energy System	Zero Crossing System
1	7.21 sec	7.798 sec
2	9.10	7.974
3	6.82	7.890
4	12.11	8.040
Total	35.24 sec	31.702 sec

$$\text{Time Ratio} = \frac{35.24}{31.702} = 1.11$$

it was busy tone. For conversations 7 of the 10 errors were audible ring tone, with 2 reorder, and 1 busy tone.

For recorded announcements the energy system required, on average, 20.7 seconds, whereas the zero crossing required 7.9 seconds. Thus the zero crossing system was about 2.6 times faster. For conversations, however, the average detection time was 8.9 seconds for the energy system, and 7.9 seconds for the zero crossing system. Therefore the zero crossing system was only about 1.1 times faster.

The only other comparison which was made between the two tone-detection systems was in terms of ease of implementation. The zero crossing system appears to be less costly to implement than the energy

system since it requires only a simple threshold device, a counter, and some simple logic, whereas the energy system requires arithmetic computations to compute the average run length, and the distance measure D_N.

VI. SUMMARY

We have presented two relatively simple methods for detecting tones based on temporal and spectral properties of the signals. The main requirement on the systems was that they be as accurate as possible in classifying tones or speech within the constraints of each individual system. Computer simulations of both systems were performed to compare and contrast the two systems. In terms of tone detection, both systems were essentially 100 percent reliable; however, the zero crossing system was from 4 to 10 times faster than the energy system. For recorded announcements the zero crossing system was more accurate than the energy system; however, for conversations the energy system was better by a small percentage. The processing time for the zero crossing system was always less than for the energy system, although the differences for speech were much less than for tones. Finally, in terms of implementation, it was argued that although both systems are relatively easy to implement, the zero crossing system is somewhat less costly than the energy system.

In summary, this study showed that the constraint of not using detailed tone frequency statistics in the detection process for the energy system led to an algorithm which was considerably slower than the zero crossing method, but equally accurate.

REFERENCES

1. L. B. Jackson, J. F. Kaiser, and H. S. McDonald, "An Approach to the Implementation of Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-16*, No. 3 (September 1968), pp. 413–421.
2. J. N. Denenberg, "Spectral Moment Estimators: A New Approach to Tone Detection," B.S.T.J., 55, No. 2 (February 1976), pp. 143–155.
3. J. J. Dubnowski, J. C. French, and L. R. Rabiner, "Tone Detection For Automatic Control of Audio Tape Drives," IEEE Trans. on Acoustics, Speech, and Signal Proc., *ASSP-24*, No. 3 (June 1976), pp. 212–215.
4. F. T. Boesch and R. E. Thomas, "A Detection Scheme for Almost Periodic Functions," unpublished work.